

Social Recommendation With Local Low Rank Matrix Approximation

Huan Zhao, Quanming Yao, Dik Lun Lee
{hzhaoaf,qyaoaa,dlee}@cse.ust.hk

Department of Computer Science and Engineering
Hong Kong University of Science and Technology
Clear Water Bay, Hong Kong

ABSTRACT

In the recommendation literature, social connections have been successfully incorporated into the traditional recommendation methods, especially the most popular state-of-art one *Matrix Factorization (MF)*. *MF* is based on the assumption that the user-item rating matrix is low rank and thus can be decomposed into two smaller matrices representing the user and item latent features. Recently, instead of assuming that the rating matrix is low rank, *Local Low Rank Matrix Approximation (LLORMA)* has been proposed based on a novel assumption that the rating matrix is composed of a set of low-rank submatrices, termed *local low rank*. Instead of factorizing the original matrix, these low-rank submatrices are factorized independently to approximate the original rating matrix. Experimental results have shown that *LLORMA* improves the recommending performance significantly comparing to standard *MF*. Unfortunately, *LLORMA* does not exploit any social information. In this paper, we propose a novel model called *Social Local Weighted Matrix Factorization (SLWMF)*, which is the first work to model social recommendation in the local low rank framework. *SLWMF* is based on the intuitive idea that the recommendation performance can surely be improved if we can intelligently integrate social connections into the local low rank framework. By assuming that there exist a set of influential users, termed *connectors*, underlying the social graph, who can influence the behaviors of people within at least three degrees, we construct the submatrices centered on a set of influential users rather than random points used in *LLORMA*. We argue that one main advantage of *SLWMF* over *LLORMA* is that the former can obtain more *meaningful* submatrices than those in the latter. Further, we also design a more effective optimizing objective functions than that of *LLORMA* and conduct extensive experiments on two real datasets. The experimental results show that *SLWMF* can further improve the recommending performance compared to several state-of-the-art baselines.

1. INTRODUCTION

Recommendation System (RS) has become a powerful tool in the big data era to deal with the information overload problem. Generally speaking, *RS* aims at helping users to get what they are interested in based on their previous behaviors. Collaborative Filtering (*CF*), a state-of-the-art method for *RS*, tries to predict users' ratings (or preferences) on unseen items based on similar users or items. For example, at Netflix, a user will give a rating (from 1 to 5) to a movie after watching. Then Netflix will predict the rating the user will give to an unseen movie according to his or her rating history. The assumption underlying *CF* is that people will be fond of the items that are similar to those they liked in the past and that similar users have similar preferences. *CF* methods can be roughly categorized into memory-based and model-based approaches. Memory-based methods mainly utilize the neighborhood information of users or items in the user-item rating matrix while model-based methods attempt to exploit systematic techniques to discover and predict users' preferences. While memory-based methods are simpler, model-based methods have stronger predicting power. Among the various model-based methods, matrix factorization (*MF*) has become the most popular one in recent years due to its good performance and scalability in large datasets [17, 16, 6, 7]. The assumption of *MF* is that the preferences of users to items are controlled by a small set of latent factors. Thus, the large user-item rating matrix can be decomposed into two smaller matrices representing user-specific and item-specific latent factors. In other words, the original matrix holds the low rank property.

Despite its good performance, *MF* still suffers from the cold-start problem and sparsity of the rating matrix. To tackle these problems, social information has been introduced into the *MF* framework based on the assumption that our behaviors and preferences are affected by our friends in the social networks. Several state-of-the-art methods have been presented in [14, 13, 5, 15, 11], most of which focus explicitly on the influences of direct friends in the social graph. For the sparsity problem, a potential direction is to split the original matrix into a set of smaller but denser submatrices, and then approximate the rating matrix out of these submatrices. Lee et al. proposes *Local Low Rank Matrix Approximation (LLORMA)*, which significantly improves the recommending performance compared to standard *MF*. In *LLORMA*, instead of assuming that the original rating matrix is low rank, termed *global low rank*, Lee et al. assumes that it is composed of a number of submatrices that hold

the low rank property, termed *local low rank*. *LLORMA* is based on the motivation that in reality there tend to be a small number of people who are interested in a small number of items, thus forming the structure of submatrices. A systematical method has been designed to construct submatrices from the rating matrix. Because the submatrices hold the low rank property, the standard *MF* framework can be easily applied to them independently. Finally, they are combined to approximate the original rating matrix.

The success of *LLORMA* demonstrates the effectiveness of the *local low rank* property. In fact, the idea of local low rank assumption has been developed in other applications and achieves state-of-the-art performances as well. For example, in colorization [22], instead of assuming the whole image is of low-rank, a group of similar patches extracted from the local area of the big image is supposed to be of low-rank. Then, in multi-label learning, a local embedding method is proposed where the low rank assumption is only imposed on subsets of similar labels [1]. The key part of the local low rank framework is how to construct a number of submatrices holding the low rank property out of the original matrix. Therefore, motivated by an intuitive idea that we can further improve the recommending performance if we can incorporate social connections intelligently into the *LLORMA* framework. In this paper, we propose a novel model called *Social Local Weighted Matrix Factorization (SLWMF)*, which is the first systematical approach to incorporate social connections into the *LLORMA* framework. Moreover, we argue that with the help of social connections, the main weakness of *LLORMA* framework can be alleviated significantly and meaningfully. Further, we observe that the main weakness of *LLORMA* lies in the construction of the submatrices. Specifically, it first randomly selects a number of user/item pair from the rating matrix, termed *anchor points*, and then chooses neighbors for the anchor points based on the user and item similarities between the neighbors and the anchor point. However, we claim that the method to randomly select anchor points is ad hoc, and hence cannot guarantee to produce meaningful submatrices.

In *SLWMF*, the submatrices are built in a novel and meaningful way. According to the well-known *Three Degrees of Influence (TDI)*,¹ we assume that there exist a set of influential users, termed *connectors*, who can exert their influences to those who are within at least three degrees to them in the social graph.² After identifying the connectors, a circle of people to whom the connector can propagate influence can be easily selected. Then, a submatrix can be constructed on a connector as well as the people in the circle and the items they interact with. To model the various strengths of the influence a connector can exert on people at different degrees, we design a heuristic scheme to gauge the strength of the influences. Further, in order to simultaneously capture the influences of direct friends in the social graph, we model them as social regularization terms which are common practice in incorporating social recommendation into standard *MF*. Then, every submatrix is approximated by a weighted matrix factorization with social regularization, termed *Local Social Regularization*. Finally, the missing predictions in the original matrix can be predicted by the weighted average of all of the submatrices containing the users and items

¹https://en.wikipedia.org/wiki/Three_degrees_of_influence

²In *TDI*, degree refers to the number of edges in the shortest path connecting two people.

concerned. We conduct extensive experiments to compare *SLWMF* with several state-of-the-art baselines and the results show that *SLWMF* can significantly improve the recommendation performance over the baselines. Besides, some interesting discoveries regarding the social theories are observed from the experiments.

In summary, the contributions of our work are as follows.

- To the best of our knowledge, *SLWMF* is the first work to systematically incorporate social connections into the state-of-the-art local low rank assumption. More importantly, *SLWMF* can capture social influences from direct friends as well as influential users (i.e., connectors).
- With the help of social connections, *SLWMF* can mitigate the main drawback of *LLORMA*, which is caused by random selection of anchor points.
- Extensive experimental results on two real datasets demonstrate the effectiveness of our model comparing to several state-of-the-art baselines, including *SoReg* [15] and *LLORMA* [9]. Besides, several interesting discoveries regarding the social theories are also figured out.

The rest of the paper is organized as follows. *MF*, social recommendation with *MF*, *LLORMA* are introduced in Section 2. We elaborate our proposed models as well as the optimization approaches in Section 3, followed by the experiments in Section 4. The experimental results, as well as the detailed analysis, are given in Section 5. Finally, we conclude our work in Section 6.

2. RELATED WORK

2.1 Matrix Factorization and Social Recommendation

Matrix Factorization (MF) has been one of the most popular approaches for rating prediction in recommendation systems. Based on the assumption that a user's rating is governed by a small number of latent factors, the rating matrix R can be approximated by the product of l -rank factors,

$$\hat{R} \approx U^T V, \quad (1)$$

where $U \in \mathbb{R}^{l \times m}$ and $V \in \mathbb{R}^{l \times n}$ with $l < \min(m, n)$, representing, respectively, the latent features of users' preferences and items. The approximating process can be completed by solving the following optimizing problem:

$$\min_{U, V} \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} (R_{ij} - U_i^T V_j)^2 + \frac{\lambda_1}{2} \|U\|_F^2 + \frac{\lambda_2}{2} \|V\|_F^2, \quad (2)$$

where $\|\cdot\|_F^2$ denotes the Frobenius norm, and I_{ij} is an indicator function, which is equal to 1 if user u_i rated item v_j and 0 otherwise. In order to avoid overfitting, two regularization terms are added to Equation 2. In the literature, this method is also termed *Regularized SVD (RegSVD)* [17].

With the proliferating of online social networks in recent years, an important trend aiming at improving recommendation performance is to incorporate social connections among users into *MF*. Most of the works are based on a well-known concept *social homophily* that people with similar interests

tend to be connected, which in turn means that users' preferences will be influenced by his friends. Ma et al. proposed a co-factorization approach based on probabilistic matrix factorization [16] to factorize the rating and social matrices simultaneously [14]. In [15], social connections have been employed as regularization terms on users' latent factors in the *MF* process, assuming that the closer two users are, the closer their latent features will be. Yang et al. proposes a circle-based social recommendation model in [21] based on the assumption that users trust different friends regarding different domains. However, as can be seen later in the paper, their notion of the circle is quite different from that in our model. In [11], Li et al. argues that users can join in different online groups because of diverse interests and overlapping community regularizations are to the users' latent feature factors in the global *MF* framework.

The uniqueness of our work is that we incorporate social connections into the local low rank assumption, rather than global low rank assumed in standard *MF*. We build our assumption based on the well-known *Three Degree Influence (TDI)* that users' preferences are affected not only by direct friends but also indirect ones, especially influential users in the social graph, named *connectors* in this paper.

2.2 LLORMA

Instead of assuming the rating matrix to be low rank, termed *global low rank*, Lee et al. argue that the rating matrix is indeed *local low rank* [9]. That is, it is composed of a set of smaller submatrices that hold the low rank property. After the locally low rank submatrices are obtained, they are factorized independently and then combined to approximate the original matrix. Thus, the model is called *Local Low Rank Matrix Approximation (LLORMA)*, which is actually motivated by the real-world fact that a small set of users are interested in a small number of items, and they form a submatrix out of the original rating matrix. Technically speaking, *LLORMA* consists of three steps. The first step is to construct the submatrices. It randomly selects q observations, termed *anchor points*, and then identifies neighbors for every anchor point based on a distance function d that measures the distance between any two observations. The set of observations selected into the submatrix with anchor point s is denoted by N_s and defined as $N_s = \{s' \in \Phi : d(s, s') < \alpha\}$, where Φ is the set of all observations in the dataset and α is a distance threshold. In the second step, a weighted factorization technique is used to solve each submatrix independently:

$$\arg \min_{U, V} \sum_{(i, j) \in N_s} w_{ij} (R_{ij} - U_i V_j^T)^2, \quad (3)$$

where w_{ij} is a weight function for every observation (i, j) in every submatrix according to the observation's similarity to the anchor point (i^*, j^*) . In the final step, the rating matrix is approximated by a linear combination of the submatrices as follows.

$$\hat{R}_{ij} = \sum_{t=1}^q \frac{w_{ij}^t}{\sum_{s=1}^q w_{ij}^s} [U_i^t (V_j^t)^T], \quad (4)$$

where w_{ij}^t is the weight of observation s_{ij} in the submatrix N_t . It means that the prediction in the original rating matrix is a weighted average of those in the submatrices. Experimental results on real datasets show that *LLORMA* improves predicting accuracy significantly comparing to *RegSVD*.

Besides *LLORMA*, several subsequent works [8, 22, 1, 12, 19] in different domains, such as image processing, topK ranking, multi-label classifications, etc., also demonstrate the effectiveness of the local low rank assumption. However, there are several disadvantages of *LLORMA*: (1) The anchor points used to construct the submatrices are chosen by randomly sampling, which does not have any reasonable justification and cannot be interpreted meaningfully; (2) The neighbors chosen for the anchor points are based on the user and item similarities between the anchor point and the observations. The similarities are calculated from the results of standard *MF*, which is less effective in practice. (3) Every submatrix in *LLORMA* is factorized independently. However, we argue that this does not achieve the goal of maximizing the prediction accuracy in the *original* rating matrix. In this paper, we adopt the local low rank assumption similar to *LLORMA*, but avoid its drawbacks with the help of social connections. For the first problem, we construct our submatrices around influential users instead of random anchor points. For the second problem, we design a heuristic scheme to measure the strength of influence a connector can exert on people at different degrees. Based on the influence strengths, we can choose the neighbors more effectively. For the last one, we design an optimization objective function directly related to the ultimate goal. Our method can also optimize all of the submatrices simultaneously. The detail of our model is elaborated in Section 3.

3. PROPOSED FRAMEWORK

3.1 Motivation

The main purpose of recommendation systems is to recommend items to users according to their preferences. However, a user's preferences can be affected by other people, in particular, influential users who have a large number of connections. As social networks proliferate, everyone in the world is connected closer to each other. This *small world phenomenon*³, also related to the famous *Six Degrees of Separation (SDS)*, implies that it is easy for people to be affected and affect other people. More importantly, it is well known that some influential users on social networks can exert large influence on a large number of people. The influential users have a large number of direct connections (e.g., a large number of friends on Facebook or followers on Twitter). According to the *TDI* theory, this phenomenon is very important for recommendation.

Based on the motivation presented above, our proposed method predicts a user's preference to an item by taking into considerations all of the influences from the circles the user belongs to as well as those from his or her direct friends. To incorporate these two social influences, extending from *LLORMA*, we propose a novel model called *Social Local Weighted Matrix Factorization (SLWMF)*, in which we assume that the original rating matrix can be represented by a set of submatrices, each of which is constructed based on a hidden circle, i.e., the users and the items they interact within the hidden circle. Moreover, as in [9], we assume that all the submatrices hold the low rank property. Thus, we can apply matrix factorization techniques to all submatrices to obtain the corresponding user's and item's latent features as in the standard *MF* framework. Then, a user's

³https://en.wikipedia.org/wiki/Small-world_experiment

rating to an item can be predicted by combining all of the predictions from the submatrices containing the user and item.

In the above, we give a general introduction of our motivation and *SLWMF* model. And in the remaining parts of this section, we will elaborate the technical details of *SLWMF*.

3.2 The *SLWMF* model

In this section, we give a formal and mathematical elaboration of our model and design the optimization approaches for it based on the *LLORMA* framework [9] and social regularization [15].

We first introduce the notations for our model. Let $\mathcal{U} = \{u_1, u_2, \dots, u_m\}$ and $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ be the sets of users and items in a recommendation system, respectively. $R \in \mathbb{R}^{m \times n}$ represents the user-item matrix, and R_{ij} represents the rating if a user u_i gives a rating to an item v_j , otherwise, $R_{ij} = 0$. Let $\mathcal{G} = (\mathcal{U}, \mathcal{E})$ represent a social network graph, where the vertex set $\mathcal{U} = \{u_1, u_2, \dots, u_m\}$ represents the users and the edge set $\mathcal{E} = \{e_1, e_2, \dots, e_p\}$ represents the social connections among all the users. The weight of every edge of \mathcal{G} is set to 1. Then, for any two users u_i, u_j , $d(u_i, u_j)$ is the distance of u_j to u_i and can be obtained from the length of the shorted path from u_i to u_j in \mathcal{G} . According to the theories of *SDS* and *TDI*, u_j is called the d^{th} degree friend of u_i and vice versa.

In our model, we assume that there are a set of circles underlying the social network, in each of which all of the users' behaviors are affected by a influential user, termed *connector*, as well as by their direct friends. We denote a circle $C_t = \{u_i | u_i \in \mathcal{U}, w^t(u_i) > 0\}$, where u^t is the connector, and $w^t(u_i)$ is an influential function for measuring the strength of influence that u^t can exert on u_i in the circle C_t . From the *TDI* theory, $w^t(u_i)$ is positively correlated with $d(u^t, u_i)$, the distance between u_i and u^t in C_t . With all the users in a circle C_t and items they interact with, we then construct a smaller rating matrix R^t . Then the original rating matrix R can be represented by a number of submatrices, $\{R^t\}_{t=1}^q$, which can share overlapping parts. Therefore, a user's rating on an item can be affected by combining influences from the connectors in multiple circles he or she belongs to. Moreover, due to the effect of *social homophily*, that people share similar interests tend to be connected, in every circle, the influences of a user's direct friends should never be ignored, which is actually the focus of previous works on social recommendation.

We then propose our model called *Social Local Weighted Matrix Factorization (SLWMF)*, which attempts to more deeply incorporate social connections with the local low rank framework to improve the recommendation performance. To capture the influences of connectors, we split the original matrix into a small number of submatrices and apply a weighted factorization technique to them, and the influences of direct friends are added as regularizations on users' latent factors in every submatrix. After obtaining all the latent factors of users and items in every submatrix, the missing rating \hat{R}_{ij} in the original rating matrix can be predicted by a weighted average of all the predictions from the submatrices:

$$\hat{R}_{ij} = \sum_{t=1}^q \frac{w_i^t}{\sum_t w_i^t} U_i^t (V_j^t)^T, \quad (5)$$

The physical meaning of this equation is that the rating of a user u_i to an item v_j is the combination effects from all of

the circles u_i belongs to. In reality, the closer u_i is to the connector u^t in circle C_t , the larger influence u^t can exert on him or her, thus the prediction $U_i^t (V_j^t)^T$ from submatrix R_t contributes more in the overall prediction.

Formally, our model consists of the following steps:

- Identify q most influential users, i.e. the connectors, in the social graph \mathcal{G} .
- Construct q submatrices centered on the q connectors.
- Design an optimization framework for factorizing all of the submatrices, $\{R^t\}_{t=1}^q$, simultaneously with local social regularization.
- Predict the missing ratings in the original matrix R by combining the predictions from all the submatrices.

3.3 Connectors Identification

The connectors are the source of influences in the circles. Thus, we need to identify the most influential users in the social network. In fact, this corresponds to a well-known *Influence Maximization* problem in social network, which attempts to find a small set of influential users in the social graph so that they can maximize the spread of influence. The *Influence Maximization* problem has been extensively in the literature. And in this paper, we adopt a heuristic method mentioned in a popular work[3], where Even-Dar and Shapira show that the best seeds for influence maximization are simply the highest degrees nodes, i.e. those with the most friends in the social graph. In specific, the set of q connectors, N_q , is obtained by the following formula:

$$N_q = \{u | u \in \mathcal{U}, f(u) \text{ is the largest } q \text{ elements}\}, \quad (6)$$

where $f(u)$ denotes the number of friends of the user u in the social network. In fact, we adopt this heuristic approach for more considerations :1) It aligns with an intuition that those with more friends tend to have more attractive characteristics thus people are more likely to adapt their behaviors or preferences to theirs; 2) In terms of the small world phenomenon, Gladwell⁴ argues that the *SDS* is dependent on few extraordinary people with large networks of friends, which he terms them as *connectors*;3) It's a heuristic but effective method, which conforms to the rule Occam's razor and the experimental results demonstrate the effectiveness of it.

After identifying the q most influential users, i.e. connectors, in the social graph \mathcal{G} , a quantitative scheme is designed to calculate the influence of a connector can exert on other users in a circle. Based on the *TDI* theory, the influence calculation is in the following:

$$w^t(u_i) = \begin{cases} f(u)/N & u_i \text{ is } u^t \\ f(u)/N \cdot e^{-\alpha \cdot d(u^t, u_i)} & d(u^t, u_i) \leq 3, \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

where $d(u^t, u_i)$ is the distance of u_i to u^t in the social graph. $f(u_i)$ denotes the number of friends of the user u_i . We also adopt a heuristic method to measure the influence of a user in the social network, i.e. $f(u_i)/N$, where $N = |\mathcal{U}|$ denotes the number of all users in \mathcal{G} . This formula also aligns with

⁴https://en.wikipedia.org/wiki/Small-world_experiment#The_social_sciences

a fact that the influence of a connector decays in an exponential manner when propagating through the social networks and the α is a factor that controls the decaying rate. In the experiments, we normalized all the influence values by the max-min normalization. One more thing to mention is that in Section 5.3, to figure the impacts of degree, we apply the formula $f(u)/N \cdot e^{-\alpha \cdot d(u^t, u_i)}$ to users within six degrees based on the SDS theory. Details are described in that section.

3.4 Submatrices Construction

To construct the submatrices centered on the q connectors, we design a method similar to the one used in the *LLORMA* [9]. In specific, for every connector, we construct the submatrix, R^t , by selecting users who are within three degrees of him or her, and items all the users including the connector interact with. Surely, there are overlapping parts of all the submatrices depending on the structure of the social graph.

3.5 Local Social Regularization

Based on the *social homophily*, we argue that the more closer u_i is to his or her direct friend, the more similar their latent factors will be in every submatrix. Therefore, similar to those used in [15], we add social regularizations on the users' latent factors in the following manner:

$$\frac{\beta}{2} \sum_{i=1}^m \sum_{f \in \mathcal{F}(i)} \text{Sim}(i, f) \|U_i - U_f\|_F^2, \quad (8)$$

where $\beta > 0$, $\mathcal{F}(i)$ is the friends set of u_i and $\text{Sim}(i, f)$ is the similarity function for measuring the similarities between u_i and u_f , which means the more similar u_i and u_j , the more closer their latent factors should be. Ma et al. incorporates these regularization terms into the global *MF* framework [15], while in our work, we integrate them into the process of submatrices factorization.

To calculate the similarities between two users, u_i, u_f , we utilize the popular Person Correlation Coefficient (*PCC*) [2]:

$$\text{Sim}(i, f) = \frac{\sum_{j \in I(i) \cap I(f)} (R_{ij} - \bar{R}_i) \cdot (R_{fj} - \bar{R}_f)}{\sqrt{\sum_{j \in I(i) \cap I(f)} (R_{ij} - \bar{R}_i)^2} \cdot \sqrt{\sum_{j \in I(i) \cap I(f)} (R_{fj} - \bar{R}_f)^2}}, \quad (9)$$

where \bar{R}_i represents the average rate of user i . Further, we employ a mapping function $g(x) = (x + 1)/2$ to bound the range of *PCC* similarity into $[0, 1]$.

3.6 Optimizing

In *LLORMA* [9], after obtaining the submatrices, Lee et al. factorized every submatrix independently. However, we argue that this is not the best option. From the prediction Equation 5, we can see that the ultimate goal is to predict the missing the ratings in the original matrix, thus it will be more proper that we optimize the factorization process regarding the original matrix, rather than every submatrix. Therefore, taking all analysis into considerations, we design

the following objective function for our *SLWMF*:

$$\begin{aligned} \min_{\{U^t, V^t\}_{t=1}^q} \mathcal{L} = & \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^n I_{ij} \left(\sum_{t=1}^q \frac{w_i^t}{\sum_t w_i^t} U_i^t (V_j^t)^T - R_{ij} \right)^2 \\ & + \frac{\beta}{2} \sum_{t=1}^q \sum_{i=1}^m \sum_{f \in \mathcal{F}(i)} \text{Sim}(i, f) \|U_i^t - U_f^t\|_F^2 \\ & + \frac{\lambda_u}{2} \sum_{t=1}^q \|U^t\|_F^2 + \frac{\lambda_v}{2} \sum_{t=1}^q \|V^t\|_F^2. \end{aligned} \quad (10)$$

The optimization problem can be solved by gradient-descent methods, which is a standard practice in *MF* framework.

4. EXPERIMENT

In this section, we conduct experiments to answer the following questions. First, what is the recommendation performance of our proposed *SLWMF* model comparing to the state-of-the-art methods? Second, how do the parameters in *SLWMF* affect the overall performance? Specifically, we want to figure out the different effects on recommendation performance of the number of the connectors and the degree used to build the circles. In the following sections, we will introduce the datasets, the evaluation measures, and the experimental settings, and then present and discuss the performance results of *SLWMF* and state-of-the-art baselines.

4.1 Datasets

We conduct our experiments on two real datasets: Yelp and Douban. The Yelp dataset is provided by the Yelp Dataset Challenge.⁵ Yelp is a location-based website where users can give ratings to and write reviews on items like restaurants, theaters, and local business centers. Douban is a Chinese website where users can rate and share their opinions on movies, books, and items. All of the ratings on the two websites are in the scale from 1 to 5. Further, on both websites, users can build Facebook-style connections to each other. Therefore, they are ideal datasets for us to evaluate the effectiveness of our model. We obtain the Yelp data from the challenge website, and the Douban Dataset is provided by the work in [15].

Since the raw data are too large and for the efficient tuning of the model, we subsample a smaller dataset from Yelp and Douban to a scale that is about the same as the well-known MovieLens-100K dataset used in the recommendation systems field. Users with too few ratings and friends are removed. In our experiments, we set the threshold to 50 for both cases. The statistics of the two final datasets used in the experiments are shown in Tables 1 and 2.

Table 1: Statistics of Ratings

	Ratings	Users	Items	Sparsity
Yelp	70479	1208	2055	2.84%
Douban	98972	1727	611	9.38%

4.2 Evaluation Metrics

We choose two evaluation metrics, namely, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), which

⁵https://www.yelp.com/dataset_challenge

Table 2: Statistics of Social Connections

Socials	Users	Edges	Sparsity
Yelp	1208	22787	3.13%
Douban	1727	21163	1.42%

are the most popular accuracy measures employed in the recommendation systems literature. *MAE* is defined as:

$$MAE = \frac{\sum_{(i,j) \in \mathcal{R}_{test}} |R_{ij} - \hat{R}_{ij}|}{|\mathcal{R}_{test}|}, \quad (11)$$

where \mathcal{R}_{test} is the set of all user-item pairs (i, j) in the test set. *RMSE* is defined as:

$$RMSE = \sqrt{\frac{\sum_{(i,j) \in \mathcal{R}_{test}} (R_{ij} - \hat{R}_{ij})^2}{|\mathcal{R}_{test}|}}. \quad (12)$$

4.3 Experimental Settings

We compare our model with several state-of-art methods:

- *RegSVD* [17]: It is the standard matrix factorization method with *L2* regularization. We use the implementation provided by Lee, et al. [10].
- *LLORMA* [9]: It is based on the local low rank assumption, which *SLWMF* also assumes. We use the implementation by Lee, et al. [10].
- *SoReg* [15]: It is a state-of-art method, which integrates social connections into the matrix factorization framework by using social connections as regularization terms. We use the implementation provided by *librec* [4].
- *SLWMF*: This is our proposed model. It (1)utilizes social connections among users to form circles for local low rank factorization and (2) performs social regularization in factorizing the submatrices. *SLWMF* is implemented based on the framework proposed by Lee, et al. [10].

We randomly split each dataset into training, validation and test data using 8:1:1 ratio. In the training process, the training data is used to fit the model while the validation data is used to obtain the best performance of every model. The test data is used to obtain the prediction errors of the models. We run every experiment five times using five random splits and show the average results. The parameter settings of *LLORMA* and *SoReg* are the same as that in the original papers [9, 15]. In all of the models, λ_u and λ_v are set to 0.001. Two ranks 10 and 20 are used in the experiments. The details of the parameters are shown in the following:

- *RegSVD*: λ_u and λ_v are set to 0.001;
- *LLORMA*: $q = 50$, $kernelWidth = 0.8$, $T = 100$ (maximum number of iterations);
- *SoReg*: The social regularization parameter $\beta = 0.01$;
- *SLWMF*: $q = 50$, $degree = 3$, $\alpha = 1$, $\beta = 0.0001$ (Local Social regularization, obtained by tuning parameters.)

5. ANALYSIS

In this section, we present and discuss the experimental results of the proposed *SLWMF* and the baselines.

5.1 Recommending Performance

Table 3 shows the performance of our proposed *SLWMF* and the baseline methods in the two datasets using two different ranks 10 and 20. The percentage of improvement of *SLWMF* over each of the baselines is shown in brackets under the respective column in the table. We can see that *SLWMF* has significant improvement over all of the baselines in both MAE and RMSE for both datasets and both ranks, while *RegSVD*, the standard *MF* method, has the worst performance, which is consistent with our expectation. Comparing to *SoReg*, which incorporates social connections as regularization terms in factorizing the global matrix, *SLWMF* makes use of *both* social circles in the local low rank framework and social regularization is submatrix factorization. Thus, *SLWMF* superior performance to *SoReg* is an evidence of the effectiveness of using social circles in the local low rank framework. Although *LLORMA* and *SLWMF* are based on the local low rank framework, *SLWMF*'s superior performance indicates that *SLWMF*'s use of social circles is better than *LLORMA*'s random anchor points in constructing the submatrices. Further, it also shows that *SLWMF*'s new optimization objective loss function is better than that of *LLORMA*. It is also interesting to note that *SoReg* obtains better performance than *LLORMA* (except in the case of Yelp at rank=20), which does not exploit social connections at all. This indicates that social connections are really important in improving recommendation accuracy and that *SoReg*'s straight-forward application of social connections as regularization terms in global matrix factorization is more effective than *LLORMA*'s local low rank framework. In the Douban dataset, we can obtain the same observations as in the Yelp dataset. Both RMSE and MAE decrease more than 5% comparing to the two stat-of-art methods. This is indeed a significant improvement.

5.2 Impact of the Number of Connectors

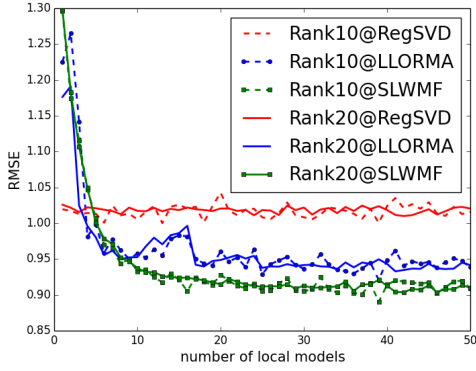
In this section, we study the impact of the number of connectors on recommendation accuracy. *SLWMF* and *LLORMA* adopt a similar framework, where connectors in *SLWMF* and anchor points in *LLORMA* play a similar role. Thus, we compare the performances of *SLWMF* and *LLORMA* by varying the number of connectors/anchor points. In both *LLORMA* and *SLWMF*, every submatrix is built centered on the connector or anchor point, which we assume holds the low rank property. In *LLORMA*, a submatrix is referred to as a local model, and every anchor point corresponds to a local model. Therefore, for consistency, we use the term "local model" to refer to both connectors and anchor points. The results are shown in the Figure 1. Because of space limitation, we only show the RMSE results because the results for both RMSE and MAE are very similar. We can see that while the RMSE of *RegSVD* is rather stable, the RMSEs of both *LLORMA* and *SLWMF* is high when the number of local models is small (1 or 2) but drop rapidly as the number of local models increases and finally stabilize to a small value. When the number of local models is larger than 10, despite some small fluctuation, the RMSEs of *LLORMA* and *SLWMF* are consistently smaller than that of *RegSVD*. This demonstrates the importance of the number of local models

Table 3: Performance Comparisons of *SLWMF* and the Baselines

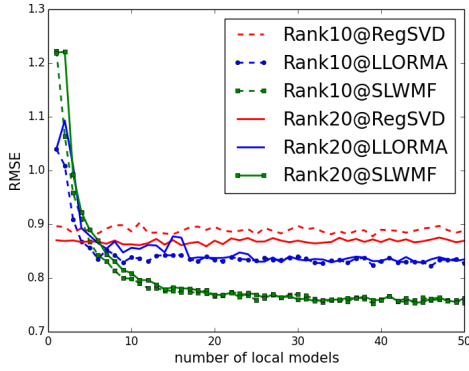
Datasets	Rank	Metrics	RegSVD	<i>LLORMA</i>	SoReg	<i>SLWMF</i>
Yelp	10	MAE	0.7929 (9.69%)	0.7348 (2.54%)	0.7232 (0.98%)	0.7161
		RMSE	1.0225 (10.17%)	0.9403 (2.32%)	0.9292 (1.15%)	0.9185
	20	MAE	0.7776 (8.92%)	0.7253 (2.36%)	0.7433 (4.72%)	0.7082
		RMSE	1.001 (9.17%)	0.9381 (3.08%)	0.9568 (4.97%)	0.9092
Douban	10	MAE	0.6848 (12.47%)	0.6625 (9.52%)	0.6333 (5.35%)	0.5994
		RMSE	0.8852 (14.34%)	0.8347 (9.15%)	0.8144 (6.89%)	0.7583
	20	MAE	0.6776 (10.86%)	0.6671 (9.46%)	0.6542 (7.67%)	0.6040
		RMSE	0.8736 (12.51%)	0.8399 (9.00%)	0.8303 (7.95%)	0.7643

for the overall performance of *LLORMA* and *SLWMF*.

Another observation is that when the number of local models is large enough, the performance remains stable. More importantly, although the trends of *LLORMA* and *SLWMF* are nearly the same, the performance of *SLWMF* is consistently and significantly better than that of *LLORMA* when the number of local models is larger than 10. In this sense, our *SLWMF* model is quite more effective.



(a) Yelp



(b) Douban

Figure 1: RMSE for Different Number of Local Models

5.3 Impact of Circle Degree on Performance

The major difference between *SLWMF* and *LLORMA* is in the construction of submatrices. Although *SLWMF* is motivated by the theory of Three Degrees of Influences, that is, users within three degrees of the connectors are selected

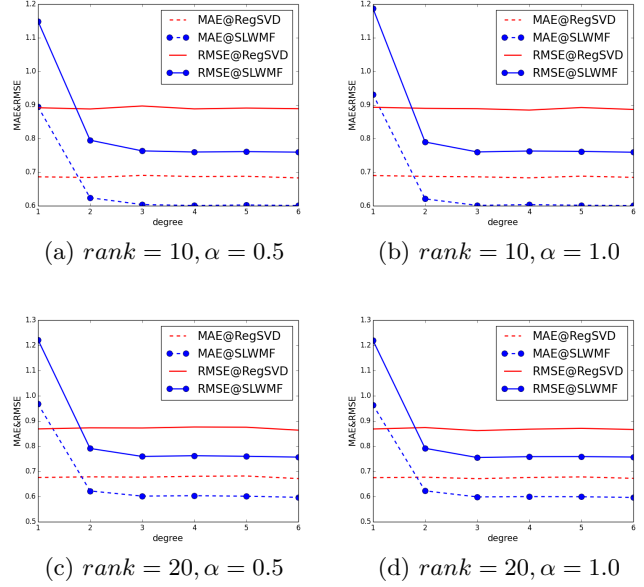


Figure 2: MAE and RMSE for Different Degrees

to construct the circles, we want to show the performance impact of the degree chosen to construct the circles.

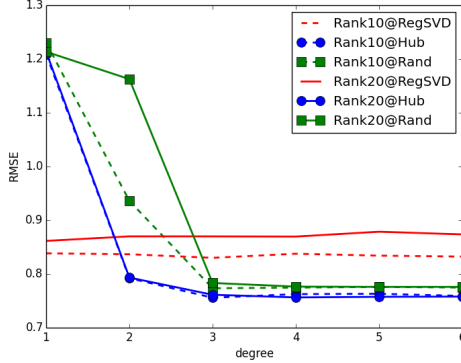
To study the effects of different degrees of the circles, we conduct experiments in multiple settings of *rank* and *α* as in Equation 7. The *rank* is set to 10 and 20 and *α* is set to 0.5 and 1.0. Thus, we get four different settings. Since the results of Douban and Yelp are very similar, we only show the result of Douban in Figure 2 because of space limitation.

As we can see, in all four figures, MAE and RMSE decrease with degree increasing. After degree is larger than three, MAE and RMSE remain nearly stable. In other words, no more improvements are gained when degree is larger than three, which aligns with the *TDI* theory. Besides, the decrease is exponential, which aligns with our assumption and weighting scheme in Equation 7. Moreover, for a particular *rank*, say, Figures 2a and 2b, the decreasing trends are very close and the stable points are both at *degree* = 3. This is an interesting discovery because it can be considered an experimental evidence of the *TDI* theory.

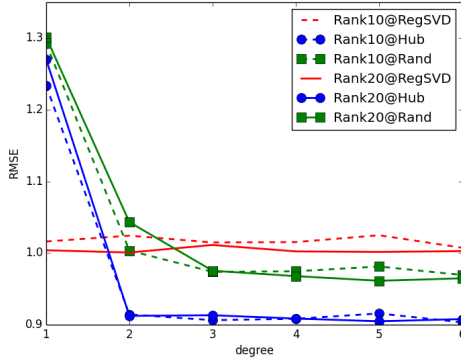
5.4 Impact Connector Identification Methods

In *SLWMF*, connectors are the key factor for constructing the submatrices. We adopt a discovery in [3] to choose the hub users as connectors in the social graph, i.e. those with

the most friends, while *LLORMA* chooses random anchor points in the training dataset for constructing the submatrices. To study the impact of these two methods of choosing connectors, we conduct experiments on both the Yelp and Douban datasets comparing the impact of the two methods on *SLWMF*'s performance. We compare the trends of RMSE by varying the degree. The *rank* is set to 10 and 20, respectively, and α is set to 1.0. The results are shown Figure 3. The two methods are denoted as *Hub* and *Rand*. As the results of RMSE and MAE are very similar, we only show the RMSE results because of space limitation.



(a) $\alpha = 1.0$, douban



(b) $\alpha = 1.0$, yelp

Figure 3: RMSE by comparing different methods of identify connectors. **Hub**:hub users; **Rand**:random users

From the results, we can see that with degree increasing, the RMSE of *Hub* is consistently lower than *Rand*, which means the performance of *Hub* is better than that of *Rand*. Besides, we can obtain an interesting observation that, before arriving the stable point, i.e. *degree* = 3, in the Douban dataset, the RMSE of *Hub* is decreasing quite faster than that of *Rand*, while on the Yelp dataset the gap of decreasing speed of RMSE is not that evident. On the contrary, after arriving the stable point, i.e. *degree* \geq 3, the performance gap of between *Hub* and *Rand* on Yelp is larger than that of Douban. These two phenomena may be attributed to the fact that the social connections among users in Yelp are much denser than that of Douban, as shown in Table 2. A sparse social network tends to mean that the influential users can propagate their influence in a large

scale more quickly. Thus, with degree increasing, RMSE decreases quickly. However, a denser networks means that after arriving the stable point the influences of influential users can propagate to a larger scale of people than those of non-influential ones, leading to the larger gap on RMSE between the *Hub* and *Rand* methods when the degree becomes large enough. In this sense, our method to identify connectors in the social graph is non-trivial comparing to the random selection in *LLORMA*.

6. CONCLUSIONS

In this paper, we explore the social recommendation with the local low rank assumption by proposing the framework *Social Local Weighted Matrix Factorization (SLWMF)*. To the best of our knowledge, *SLWMF* is the first work to systematically integrate the social connections with the effective *LLORMA* framework. On the one hand, *SLWMF* further improves the recommending performance compared to the state-of-art social recommendation approaches, most of which are conducted in the *global low rank assumption*. On the other hand, with the help of social connections, *SLWMF* alleviates the main drawback of *LLORMA* according to the well-known *TDI* theory. In specific, we propose a more effective and meaningful method for constructing the submatrices as well as a more accurate optimization objective function used in submatrix factorization. We then conduct extensive experiments on two real datasets, and the results demonstrate that our *SLWMF* method can significantly decrease the RMSE and MAE recommendation accuracy, which implies the effectiveness of our assumptions and model.

In the future research work, we plan to study more sophisticated methods like PageRank for identifying connectors and explore their impact. Since the submatrices in *SLWMF* are built centered on the influential users, if we can identify more accurate influential users, it is very likely that the recommending performance will be improved. Further, the connections in the *SLWMF* are bi-directional like Facebook-style friends. We want to utilize uni-directional connections, like Twitter's following connections, or even the signed social networks.

7. REFERENCES

- [1] K. Bhatia, H. Jain, P. Kar, M. Varma, and P. Jain. Sparse local embeddings for extreme multi-label classification. In *Advances in Neural Information Processing Systems*, pages 730–738, 2015.
- [2] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.
- [3] E. Even-Dar and A. Shapira. A note on maximizing the spread of influence in social networks. In *International Workshop on Web and Internet Economics*, pages 281–286. Springer, 2007.
- [4] G. Guo, J. Zhang, Z. Sun, and N. Yorke-Smith. Librec: A java library for recommender systems. In *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 23rd International Conference on User Modeling, Adaptation and Personalization*, 2015.
- [5] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 135–142. ACM, 2010.
- [6] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.
- [7] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. Number 8, pages 30–37. IEEE, 2009.
- [8] J. Lee, S. Bengio, S. Kim, G. Lebanon, and Y. Singer. Local collaborative ranking. In *Proceedings of the 23rd international conference on World wide web*, pages 85–96. ACM, 2014.
- [9] J. Lee, S. Kim, G. Lebanon, and Y. Singer. Local low-rank matrix approximation. In *Proceedings of The 30th International Conference on Machine Learning*, pages 82–90, 2013.
- [10] J. Lee, M. Sun, and G. Lebanon. Prea: Personalized recommendation algorithms toolkit. *Journal of Machine Learning Research*, 13(Sep):2699–2703, 2012.
- [11] H. Li, D. Wu, W. Tang, and N. Mamoulis. Overlapping community regularization for rating prediction in social recommender systems. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 27–34. ACM, 2015.
- [12] X. Li, B. Shen, B.-D. Liu, and Y.-J. Zhang. A locality sensitive low-rank model for image tag completion. *IEEE Transactions on Multimedia*, 18(3):474–483, 2016.
- [13] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval*, pages 203–210. ACM, 2009.
- [14] H. Ma, H. Yang, M. R. Lyu, and I. King. Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940. ACM, 2008.
- [15] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296. ACM, 2011.
- [16] A. Mnih and R. Salakhutdinov. Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264, 2007.
- [17] A. Paterek. Improving regularized singular value decomposition for collaborative filtering. In *Proceedings of KDD cup and workshop*, volume 2007, pages 5–8, 2007.
- [18] B. Recht and C. Ré. Parallel stochastic gradient algorithms for large-scale matrix completion. *Mathematical Programming Computation*, 5(2):201–226, 2013.
- [19] K. Wang, X. Duan, J. Ma, C. Sha, X. Wang, and A. Zhou. Local weighted matrix factorization for implicit feedback datasets. In *Database Systems for Advanced Applications*, pages 381–395. Springer, 2016.
- [20] J. Yang and J. Leskovec. Overlapping community detection at scale: a nonnegative matrix factorization approach. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 587–596. ACM, 2013.
- [21] X. Yang, H. Steck, and Y. Liu. Circle-based recommendation in online social networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1267–1275. ACM, 2012.
- [22] Q. Yao and J. T. Kwok. Colorization by patch-based local low-rank matrix completion. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, pages 1959–1965, 2015.
- [23] H.-F. Yu, C.-J. Hsieh, S. Si, and I. Dhillon. Scalable coordinate descent approaches to parallel matrix factorization for recommender systems. In *International Conference on Data Mining*, pages 765–774, 2012.